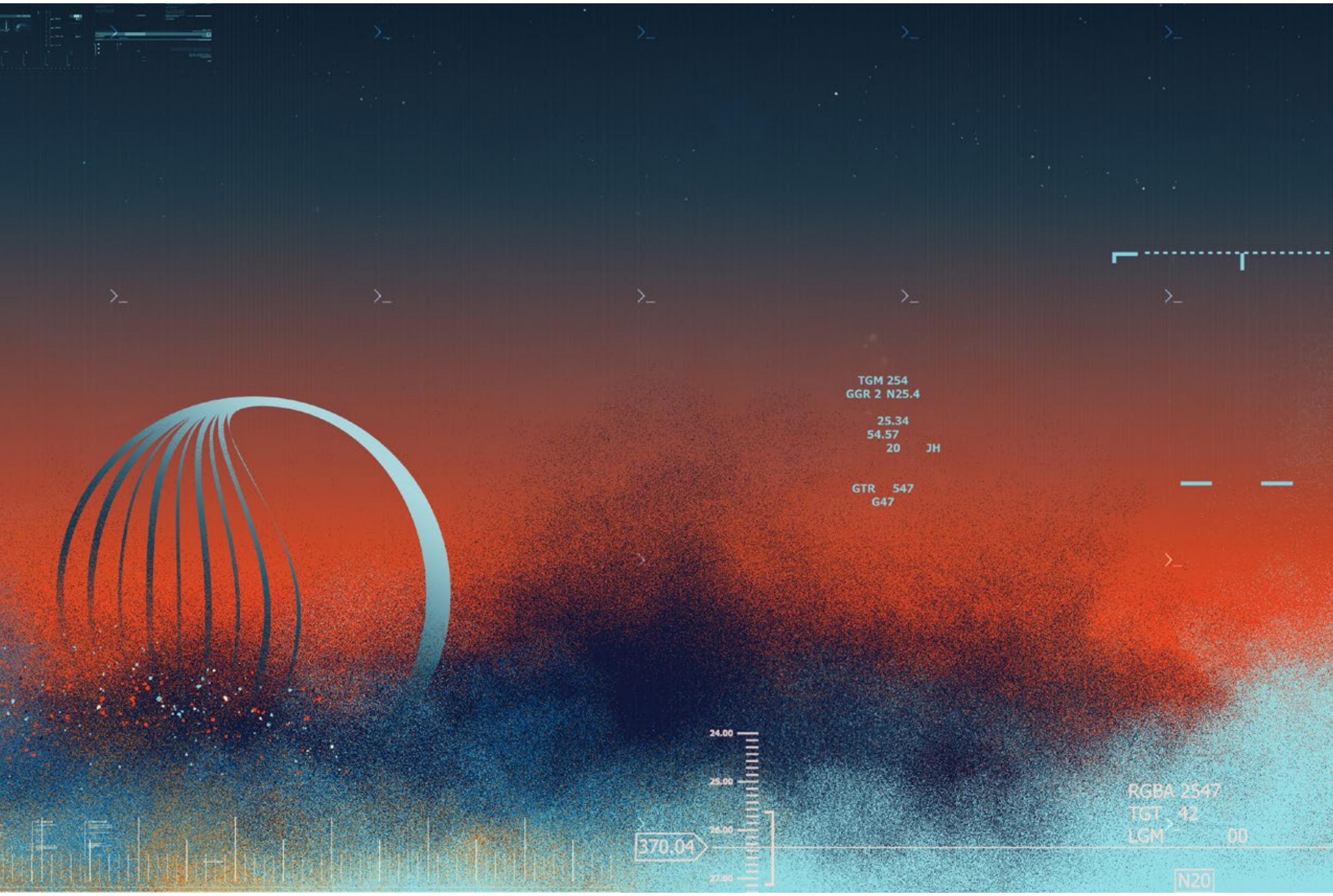
signal65



Digital Twins as the Control Plane for Physical AI

Fleet-Scale Deployment with Dell AI Infrastructure

AUTHOR

Brian Martin

AI Data Center Performance | Signal65

MARCH 2026

IN PARTNERSHIP WITH

DELL Technologies

550 AERO

Executive Summary

Every organization operating physical assets, from fleets to factories and energy infrastructure to industrial equipment, faces the same structural problem. The systems generating the most operational risk are the hardest to train AI on, because failures are rare by design. The result is a gap between the data organizations have (normal operations) and the intelligence they need (early warning of what hasn't happened yet). Closing that gap has traditionally required waiting for failures to accumulate in the real world or accepting the limitations of physics models that don't capture how individual assets actually behave.

This paper presents a different way of thinking about what a digital twin is. Not a simulation layer. Not a visualization tool. A continuously learning operational mirror of a physical asset. One that ingests real telemetry, learns the behavioral signature of that specific asset, generates scenarios the physical world cannot safely produce, and becomes the reference against which real-world performance is continuously judged. The physical asset becomes the sensor. The digital twin becomes the operational authority.

Leveraging Dell Technologies PowerEdge servers and PowerSwitch networking, 550 Aero demonstrated this architecture in one of the most demanding environments possible, aircraft operations, using real flight telemetry comprising millions of data points from six aircraft platforms, with detailed validation results presented for two. The findings are directly transferable to any domain where safety-critical physical systems generate time-series telemetry: industrial equipment, energy infrastructure, autonomous vehicles, maritime assets. The same closed loop applies everywhere: real telemetry trains the twin > the twin generates operational scenarios > deviations from the twin surface anomalies > the twin improves as the asset operates. A key secondary discovery: the training process itself became a sensor health diagnostic, identifying a failing component in active service before conventional monitoring flagged it for repair.

The infrastructure architecture follows the same logic as the business model: Train Big, Deploy Smart. The Dell PowerEdge XE7745 is the factory, GPU-dense to train per-asset digital twins at fleet scale. The Dell PowerEdge R770 is the production line, running continuous real-time inference across an entire monitored fleet, scaling to thousands of assets in a single rack. Enterprise operational intelligence for physical systems does not require exotic infrastructure at every layer. It requires the right hardware at each phase of the digital twin lifecycle and a platform architecture that makes that lifecycle repeatable across any asset class.

Key Highlights

Reduce Safety Risk
Identify and address emerging performance deviations before they escalate into safety incidents or operational failures

Increase Asset Uptime
Shift from reactive fixes to predictive intervention, keeping critical systems available and operating at peak performance

Lower Maintenance Cost
Prioritize maintenance based on actual asset behavior, reducing unnecessary inspections, avoiding failures, and extending component life

Synthetic Data Imperative for Safety-Critical AI

The Data Scarcity Challenge in Aviation

AI models are only as good as the data they train on, and aviation presents a paradox: the industry's exceptional safety record means the failure events that anomaly detection systems must recognize are extraordinarily rare in operational data. Fleet-wide monitoring generates vast volumes of normal-operations telemetry, but edge cases like slow-developing faults, unusual degradation patterns, and compound failures are precisely what models need and what the data lacks.

The problem goes deeper than rare anomalies. Real-world aircraft training data is fundamentally constrained by the cost and safety boundaries of the physical world. For a newly certified engine or a recently delivered aircraft, available data may consist of nothing more than engine test stand runs and a handful of initial flights. The data that does exist is heavily biased toward the most common operating regime: cruise-phase flight at constant power settings. A typical 3-hour cross-country flight likely contains over two hours of nearly identical cruise telemetry, valuable for establishing baseline behavior, but offering limited variability for training ML models that must detect anomalies across the full flight envelope.

Traditional approaches to this problem include physics-based simulation, hand-crafted fault injection, and rule-based threshold alerting. Each has fundamental limitations: physics models are expensive to build and may not capture real-world variability, fault injection is constrained by known failure modes, and threshold alerting is reactive by design. What is needed is a generative approach that can learn the statistical structure of real sensor behavior and extrapolate into plausible anomaly scenarios that have not yet been observed.

The Anomaly You Cannot Train On

Consider the following real-world scenario involving a slow-developing oil system degradation in a piston engine aircraft. The pattern unfolds over hours: oil temperature rises incrementally, oil pressure drops in small steps correlated with power settings, and cylinder head temperatures begin trending upward as lubrication degrades. At any single point in the progression, every parameter remains within published normal ranges. A threshold-based system sees nothing. A pattern-based system needs training examples, but this exact scenario may appear in real fleet data only once per 100,000 flight hours.

Models trained on large numbers of similar assets, such as engines, motors, or pumps, can only represent the average or overall range of behaviors. These models cannot see a given asset is operating outside its "normal" operating parameters. A successful digital twin needs to be trained on and model a single specific asset.

Generating Synthetic Time-Series Data

Diffusion, or generative denoising, models offer specific advantages over alternative generative approaches for time-series synthesis. Unlike GANs (Generative Adversarial Networks), training is stable with no mode collapse, ensuring the generated data covers the full range of real operating conditions rather than collapsing to a narrow subset. Unlike VAEs (Variational Autoencoders), the iterative denoising process preserves fine-grained temporal detail, critical when subtle inter-parameter correlations carry diagnostic meaning, such as the relationship between manifold pressure and EGT (Exhaust Gas Temperature) spread. The conditional generation framework naturally accommodates flight-phase-dependent behavior: the same model generates both idle-on-ramp and cruise-at-altitude telemetry based on input conditions.

550 Aero leverages Diffusion-TS (diffusion time-series) models to directly address the data scarcity problem by training a digital twin on each aircraft's real normal-operations data, then using that twin to generate

hundreds of scenarios, like the oil degradation scenario, synthetically – with all the physics, the noise, and the aircraft-specific behavior patterns intact.

The Diffusion-TS Digital Twin Architecture

From Fleet Data to Specific Aircraft Digital Twins

The 550 Aero pipeline constructs one diffusion model per asset (aircraft). Each model learns the mapping from flight condition inputs to engine/electrical system outputs for that specific aircraft, capturing its individual wear patterns, manufacturing variance, and operating characteristics

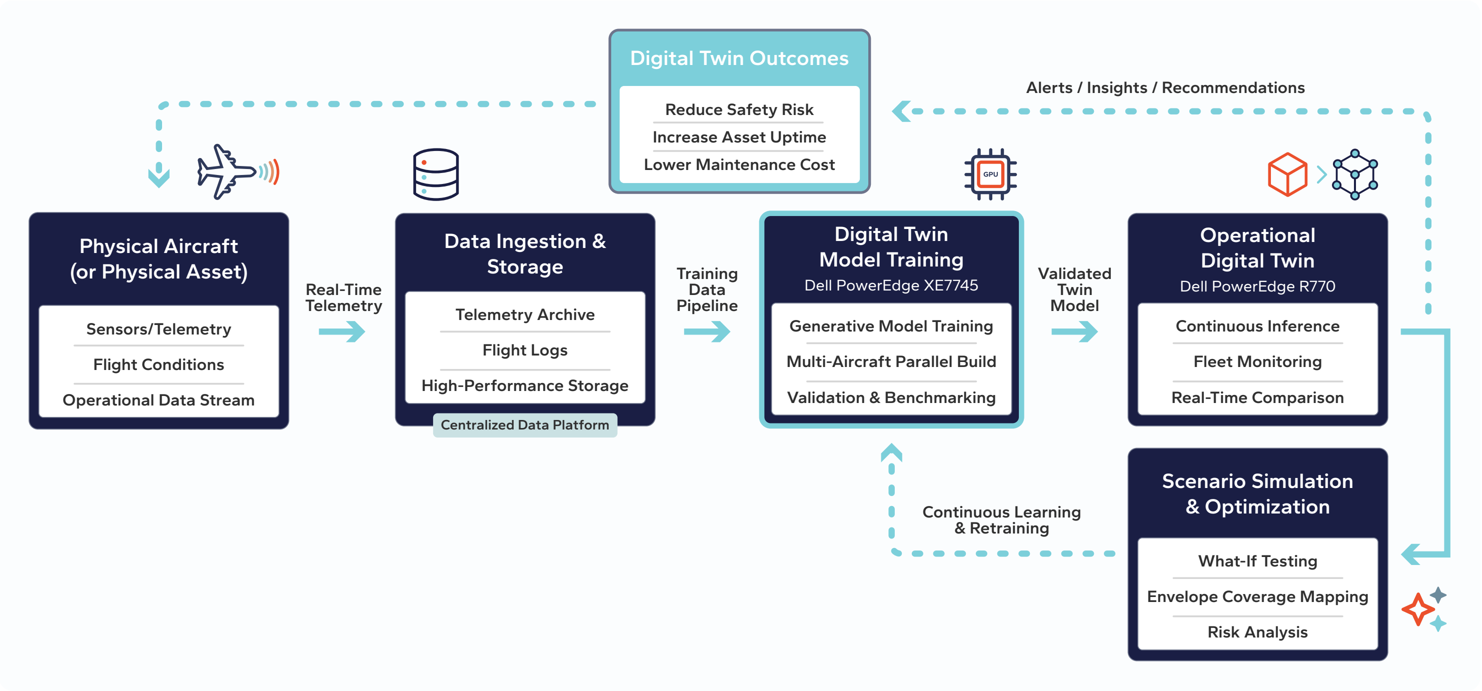


Figure 1: *Architectural Overview*

Implemented Input/Output Architecture:

Aircraft	Inputs (Conditioning)	Outputs (Generated)
AC-001: Certified Piston Single (Garmin G1000)	6 Channels: Altitude, Outside Air Temperature, Air Speed, Fuel Flow, MAP (Manifold Pressure), RPM	16 Channels: Volts, Amps, Oil Temperature, Oil Pressure, 6×EGT, 6×CHT (Cylinder Head Temp)
AC-002: Experimental Piston Single (Dynon SkyView HDX)	9 channels: Air Speed, Pressure Altitude, Vertical Speed, Outside Air Temp, TAS, Density Altitude, RPM, MAP, Fuel Flow	19 channels: 6×CHT, 6×EGT, Oil Pressure, Oil Temp, Fuel Pressure, Volts, Amps, RPM, Fuel Flow

The architectural decision to build per-aircraft models rather than fleet-wide models is driven by the observation that individual aircraft exhibit measurably different sensor response characteristics even within the same type. AC-001 (certified piston single) and AC-002 (experimental piston single) use entirely different avionics systems, sensor suites, and data formats, yet the same architecture serves both, demonstrating platform generalizability.

Training Data Volume (AC-001):

- 79 real flight log files from Garmin G1000 avionics , representing 116 total hours with mean flight time of 134 minutes
- 418,000+ rows of 1Hz engine telemetry
- Flights from multiple airports across the southwestern United States
- Spanning approximately 5 months of operational flying

Diffusion-TS Model Architecture

The 550 Aero digital twin uses a Transformer-based denoising network within a DDPM framework. The model learns to predict the noise added during the forward diffusion process, enabling iterative denoising from pure Gaussian noise back to structured time-series output, conditioned on flight state inputs.

Implemented Model Specifications

Component	Details
Architecture	Transformer Denoiser: Transformer encoder with epsilon parameterization
Parameters	5M trainable parameters
Encoder Layers	4-layer Transformer encoder, pre-norm (LayerNorm before attention)
Attention Heads	8 heads, d_model = 256, dim_feedforward = 1024
Timestep Embedding	Sinusoidal positional encoding + 2-layer MLP (SiLU activation)
Diffusion Schedule	Cosine beta schedule (Nichol & Dhariwal 2021), T = 1,000 steps
Input Projection	Linear(target_dim > d_model) + Linear(cond_dim > d_model), additive fusion
Normalization	MinMaxScaler to [-1, 1] range, fitted on training data only
Loss Function	MSE between predicted noise and actual noise
Epochs	200 with best-model checkpointing on validation loss

X-Plane Synthetic Input Generation

A critical component of the pipeline is X-Plane 12 integration. X-Plane provides a high-fidelity flight dynamics simulation that generates physically accurate flight condition inputs (altitude, airspeed, OAT, manifold pressure, RPM, fuel flow), but its engine thermodynamics model is simplified. X-Plane outputs zero for most individual CHT channels and uses generic EGT values. The 550 Aero digital twin fills this gap: X-Plane flight conditions are fed as conditioning inputs, and the diffusion model generates realistic per-cylinder engine telemetry that the physics engine cannot produce.

The Synthetic Data Advantage: Breaking Free from Physical-World Constraints

REAL DATA LIMITATIONS:

- Rigid and bounded by cost/safety: cannot fly into oil starvation to collect training data
- New aircraft/engines have minimal data: may have only test stand runs or a handful of flights
- 60–70% of a typical flight is constant-power cruise telemetry with limited variability for ML training
- Environmental conditions are not controllable; flight envelope coverage is limited

SYNTHETIC DATA CAPABILITIES (X-Plane + Diffusion Model):

- Generate unlimited flight-phase variability: takeoff, climb, cruise, descent, go-around
- Produce abnormal corner cases safely: oil degradation, CHT exceedances, electrical system fades
- Control environmental variables precisely: OAT, altitude, density altitude, humidity combinations
- Create what-if scenarios outside normal operating windows: over-temp, over-boost, extreme lean/rich

Initial Inference Results

During stabilized cruise conditions, the model achieves CHT predictions within 2–5% of actuals across all 6 cylinders, EGT predictions with 1–3%, voltage <0.1%, oil pressure within 3-4%, and oil temperature within 5–6%.

When inferenced on X-Plane flight control inputs, the diffusion model generates physically plausible engine telemetry across all metrics: voltages at ~27.7V (realistic for a 28V alternator), oil temperature ~160–165°F, oil pressure scaling with RPM from ~33 PSI at idle to ~47 PSI at cruise, CHTs in the ~185–195°F range with per-cylinder variation, and EGTs at ~850–950°F at idle rising to ~1,050–1,100°F at cruise power. Critically, X-Plane outputs zero for CHT2–CHT6 and EGT2–EGT6, while the diffusion model generates differentiated per-cylinder values, demonstrating the digital twin is generating realistic engine behavior from flight condition inputs alone, not copying simulator outputs.

Dell Infrastructure: Train Big, Deploy Smart

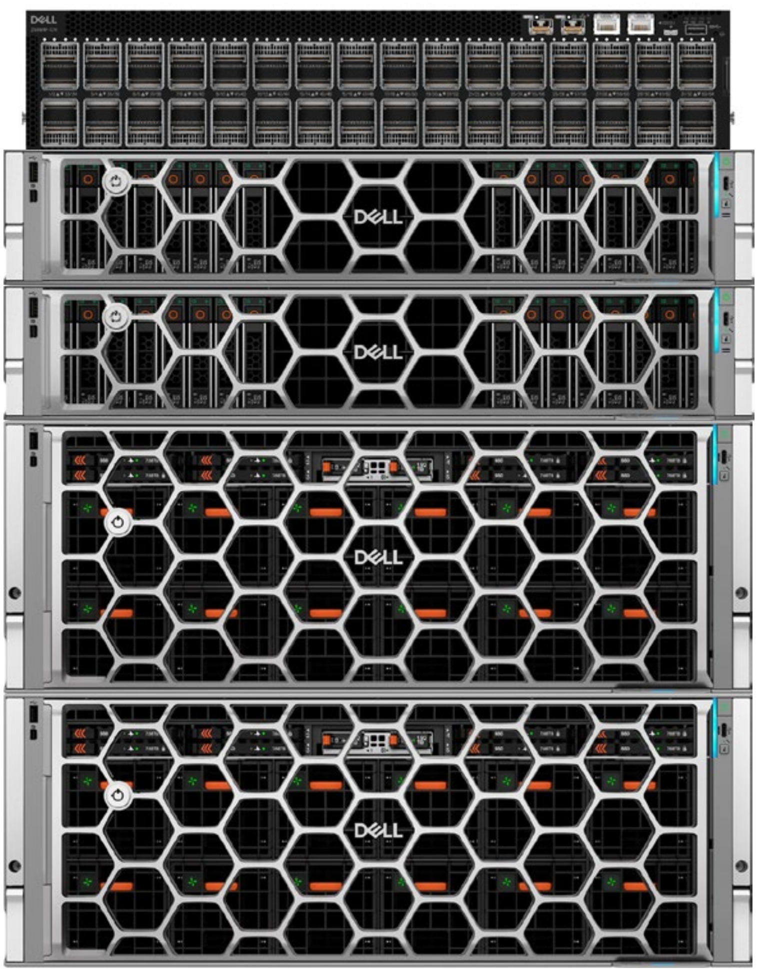
Diffusion model training is among the most compute-intensive workloads in generative AI. Each training step requires a forward diffusion pass, a denoising prediction, and a backward pass through a Transformer with 1,000 potential timestep values. Training aircraft-specific digital twins across a fleet of hundreds of aircraft requires enterprise-grade GPU compute, high-throughput storage and networks, and fleet-scale orchestration.

But training is only half the story. Once digital twins are built, they must serve production inference continuously, processing incoming telemetry, flagging anomalies in near real-time, and scaling across an entire monitored fleet. These two requirements map to two distinct infrastructure profiles, and Dell's PowerEdge portfolio addresses both.

Train Big, Deploy Smart

- Dell PowerEdge XE7745 GPU-dense training powerhouse: build per-aircraft digital twins at fleet scale
- Dell PowerEdge R770 storage-rich inference workhorse: serve a 200-aircraft fleet from 4 nodes
- Train once per tail number on XE7745 > deploy trained checkpoints to R770 for continuous operations

Testing Configuration



- Dell Z9864F-ON
- SONIC 4.4.0

Cluster:

- 2 Dell PowerEdge R770
- Dual Intel Xeon 6760P
- 2 TB DDR5 Memory
- Dual PERC H975i
- 2 NVIDIA L40S
- 2 Broadcom BCM57608 400GbE RoCEv2 Network Controller
- Ubuntu 24.04 LTS
- CUDA 13.0 / Driver 580.95.05

Cluster:

- 2 Dell PowerEdge XE7745
- Dual AMD EPYC 9555
- 2.3 TB DDR5 Memory
- 8 NVIDIA RTX Pro 6000 Blackwell Server
- 8 Broadcom BCM57608 400GbE RoCEv2 Network Controller
- Ubuntu 24.04 LTS
- CUDA 13.0 / Driver 580.95.05

Figure 2: *Dell Training and Inference Clusters*

Training on Dell PowerEdge XE7745

The Dell PowerEdge XE7745 is purpose-built for the compute-intensive phase of the digital twin lifecycle. Each per-aircraft model requires sustained GPU throughput over hundreds of training epochs, processing tens of thousands of 60-second windows through 4-layer Transformer attention with 1,000 diffusion timesteps per sample.

XE7745 Training Environment

Component	Details
Server	Dell PowerEdge XE7745 (4U air-cooled chassis)
CPU	Dual AMD EPYC 9555 Series up to 192 cores per socket, Zen5 architecture
GPU / Accelerator	Up to 8× double-wide PCIe accelerators (600W each) such as NVIDIA H200 or RTX Pro 6000 Blackwell
System Memory	Up to 3TB DDR5 (24 DIMM slots, 6400 MT/s, 2.3TB as tested)
Storage	Up to 8× E3.S Gen5 NVMe SSDs (~122TB max) for high-throughput dataset access
Networking	8× PCIe Gen 5.0 slots + OCP 3.0 Ethernet for multi-node scaling
Power	3200W Titanium efficiency, hot-swap redundant PSUs
Management	iDRAC10 with Silicon Root of Trust

Why the XE7745 for Training:

- GPU density: 8× 600W accelerators in a single 4U chassis eliminates multi-node complexity for per-aircraft training
- Memory bandwidth: DDR5 at 6400 MT/s with 192-core EPYC keeps the data pipeline saturated during training
- NVMe throughput: Gen5 storage eliminates I/O bottlenecks when loading flight logs from hundreds of aircraft

Inference on Dell PowerEdge R770

Once a digital twin is trained, the trained model checkpoint (~38MB .pt file) transfers to the deployment platform for production inference. This is where the Dell PowerEdge R770 enters the architecture as the right tool for the deployment phase. Inference against a trained diffusion model is orders of magnitude less compute-intensive than training: the backward pass is gone, gradient memory is freed, and each aircraft's inference runs independently. The R770 handles this workload with room to spare, at a fraction of the XE7745's cost and rack footprint.

R770 Deployment Environment

Component	Details
Server	Dell PowerEdge R770 (2U rack server)
CPU	Dual Intel Xeon 6 processors - up to 86 P-cores or up to 144 E-cores
GPU / Accelerator	Two double-slot GPUs such as NVIDIA H200 or RTX Pro 6000 Blackwell or 6 single-slot GPUs
System Memory	Up to 8TB DDR5 (32 DIMM slots, 6400 MT/s)
Storage	OCP DC-MHS compliant 1GbE to 400GbE options for fleet telemetry ingestion
Networking	Multiple PCIe Gen 5.0 and OCP 3.0 Ethernet
Power	800W to 3200W PSU options (Platinum and Titanium efficiency)
Management	iDRAC10 with Silicon Root of Trust

The R770's value proposition centers on continuous fleet-wide inference: processing incoming 1Hz telemetry from every monitored aircraft, running windowed diffusion inference at 30-second stride, and delivering anomaly flags within seconds of data receipt. The workload is extremely parallel across aircraft, each digital twin processes its own aircraft's data stream independently, with no cross-aircraft computation. This maps perfectly to the R770's multi-GPU configuration: each GPU serves a batch of aircraft simultaneously.

Fleet Scale-Out

OPERATOR SCALE: 1 R770 node handles a 50-aircraft fleet in real-time

>> Purpose: Owner-operated fleets, charter operators, flying clubs with shared AI monitoring

REGIONAL SCALE: 4 R770 nodes handle a 200-aircraft regional fleet

>> Purpose: Regional airlines, fractional ownership operators, FBO networks

ENTERPRISE SCALE: 20 R770 nodes handle a 1,000-aircraft major carrier fleet

>> Purpose: National carriers, leasing companies, MRO networks with fleet-wide monitoring

Based on measured per-aircraft inference throughput of 38ms on L40S

How Much Data is Required

A critical step in AI model development and deployment is knowing how much data is required to achieve training convergence without overfitting and to meet inference accuracy targets.

A systematic data ablation study was conducted on both aircraft platforms to answer a critical deployment question: how much real flight data is needed to train a functionally useful digital twin? Each tier uses a nested subset of training data (5% > 10% > 25% > 50% > 75% > 100%), ensuring smaller tiers are strict subsets of larger tiers.

- AC-001 (Garmin G1000): 6-tier ablation over 79 total training flights. Holdout: 1 long-duration flight (13,738 rows, 99th percentile by length). 200 epochs per tier.
- AC-002 (Dynon SkyView HDX): 6-tier ablation over 37 valid training sessions. Holdout: Session 57 (5,434 rows, 87th percentile). 200 epochs per tier.

Synthetic Data Fidelity

Fidelity at Full Training Dataset in Mean Absolute Percentage Error (MAPE)

Fidelity Metric	AC-001 (Garmin G1000)	AC-002 (Dynon HDX)
Cruise-phase EGT accuracy (6-cyl avg)	0.6% MAPE	1.9% MAPE
Cruise-phase CHT accuracy (6-cyl avg)	2.6% MAPE	9.4% MAPE
Voltage tracking accuracy	<0.1% MAPE	0.1% MAPE
Oil pressure accuracy	3.3% MAPE	3.4% MAPE
Oil temperature accuracy	3.0% MAPE	17.4% MAPE †
Overall engine health MAPE † †	1.69%	6.3%

† AC-002 oil temp: anomalously high due to failing sensor discovered during ablation analysis.

†† Overall engine health MAPE excludes the Amps channel (pilot behavior, not engine health). Amps reflect avionics load and lighting and not a reliable engine health signal.

Data Efficiency Curve

The ablation study's primary contribution is an empirical answer to a question every fleet operator faces: how many flights does it take to build a useful digital twin for a new aircraft?

Training Data Efficiency

Aircraft	5% Dataset	25% Dataset	75% Dataset	Full Dataset	Elbow
AC-001 (Garmin G1000)	6.4% (4 flights)	1.9% (20 flights)	1.68% (60 flights)	1.69% (79 flights)	~20 flights
AC-002 (Dynon HDX)	24.2% (2 flights)	10.3% (10 flights)	8.5% (28 flights)	6.3% (37 flights)	~18 flights

All MAPE values exclude Amps (pilot behavior, not engine health)

Both aircraft independently show a sharp improvement around 20 flights, a convergence that holds despite different avionics systems, data formats, and amounts of available training data (79 flights vs. 37 flights). For AC-001, the 25% tier delivers 1.9% overall MAPE and the model reaches saturation by 60 flights: the 75% tier (1.68%) and 100% tier (1.69%) are statistically identical.

Metric Predictability Hierarchy

Tier	Metrics	Physical Explanation	Elbow
Easy (sub-2% at 25%)	EGTs, RPM, Fuel Flow, Volts	Physically deterministic: directly follows throttle, mixture, electrical load	~10 flights
Medium (2–10% at 25%)	CHTs, Oil Pressure, Fuel Pressure	Thermal lag, cylinder variation, nonlinear pressure-temperature relationships	~20 flights
Hard (rarely converges)	Oil Temperature, Amps	Oil temp: ambient-dependent, large thermal mass; Amps: pilot behavior, not engine state	50+ flights

Sensor Quality vs. Data Volume

AC-001 (factory-calibrated Garmin G1000) achieves 1.69% overall engine health MAPE at 79 flights. AC-002 (Dynon SkyView HDX, hand-assembled experimental aircraft) achieves 6.3% with 37 training sessions. This difference likely reflects sensor calibration quality, data consistency, and avionics logging precision. Sensor data quality is at least as important as raw flight count for model accuracy.

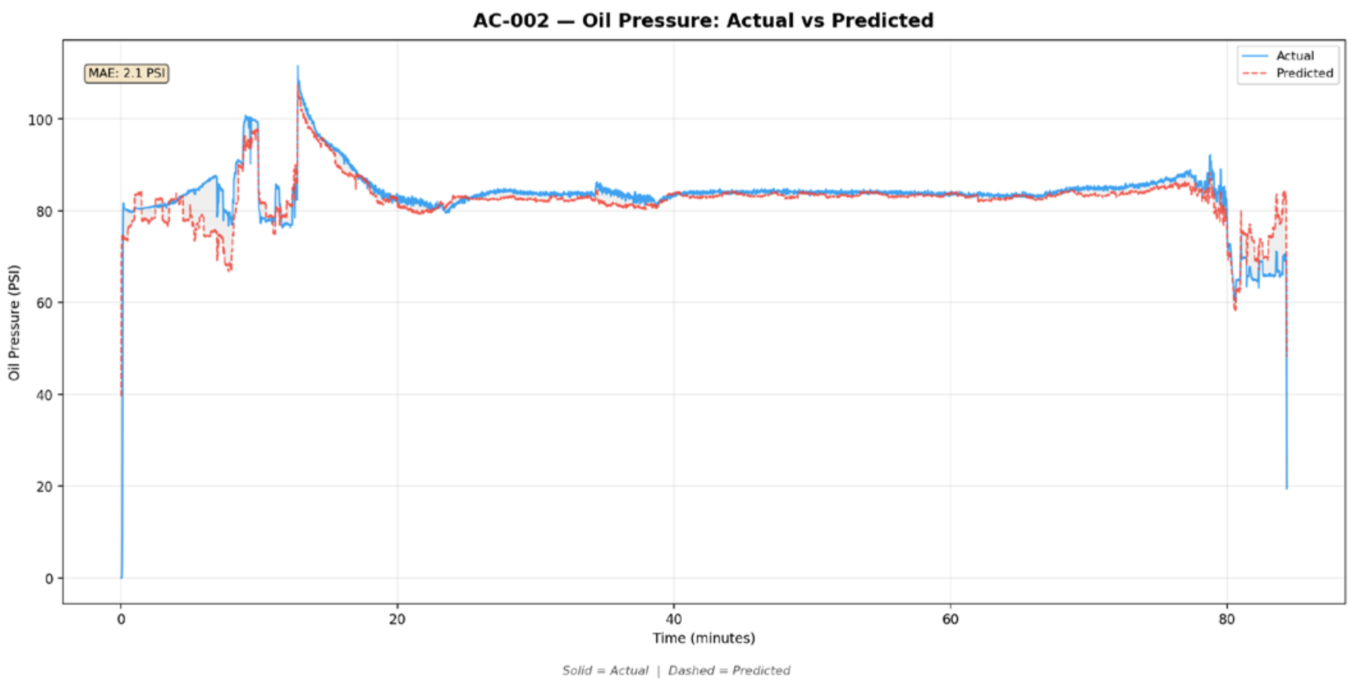
Holdout Representativeness

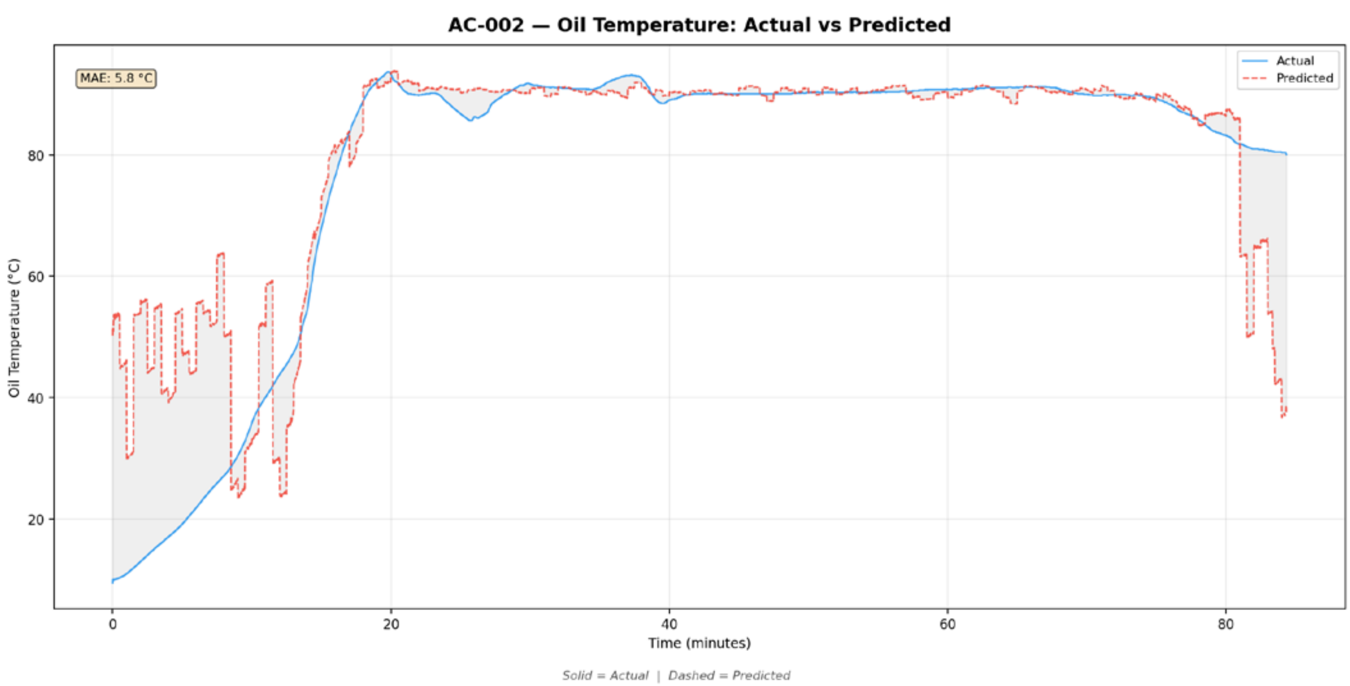
The AC-001 holdout flight is a 99th-percentile outlier (13,738 rows, approximately 4 times the median training flight duration), representing a harder generalization challenge that makes MAPE appear worse than it would for a representative holdout. Holdout selection methodology significantly impacts reported accuracy and should be standardized when comparing models across aircraft.

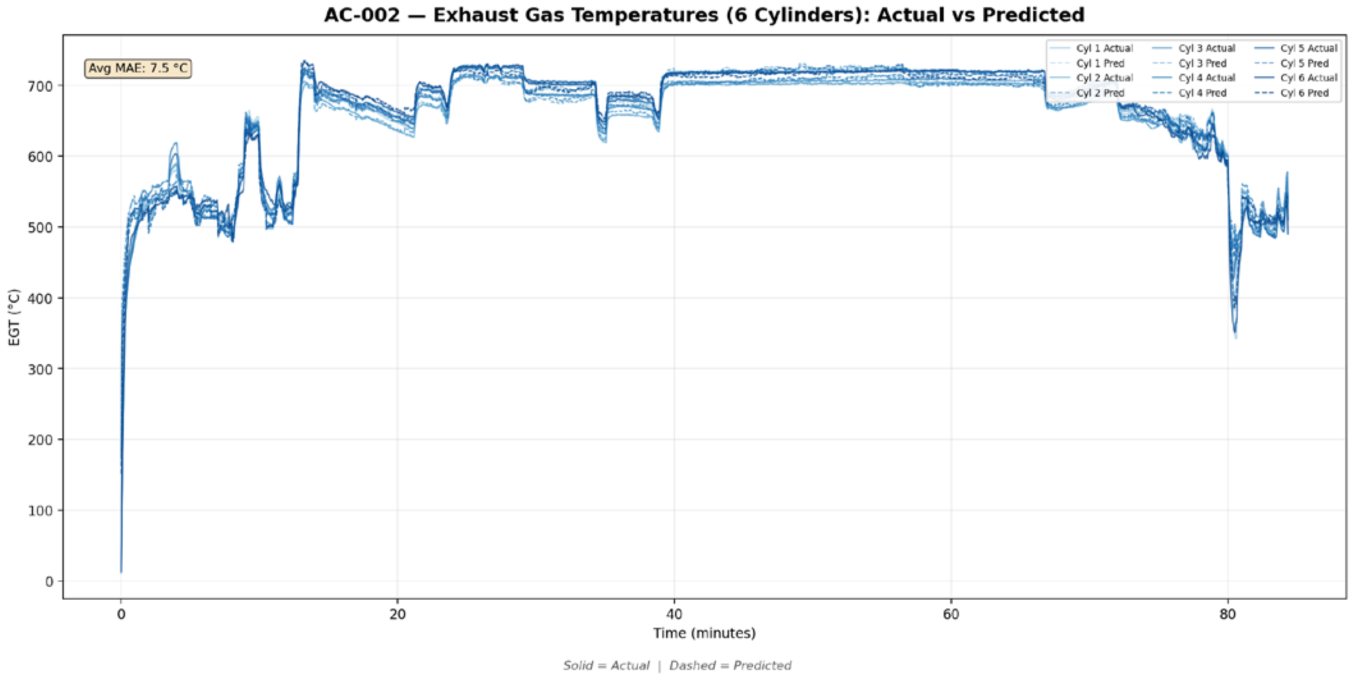
Model Saturation

For AC-001, the model fully saturates at 60 flights: MAPE at the 75% tier (1.68%) and 100% tier (1.69%) are statistically identical. Adding 19 more flights produced no measurable improvement. This saturation point defines the practical ceiling. To push accuracy further would require higher-quality sensor data, a longer context window, or architectural changes.

Figure 3, below, shows diffusion model predictions relative to actual values across four critical metrics for AC-001.







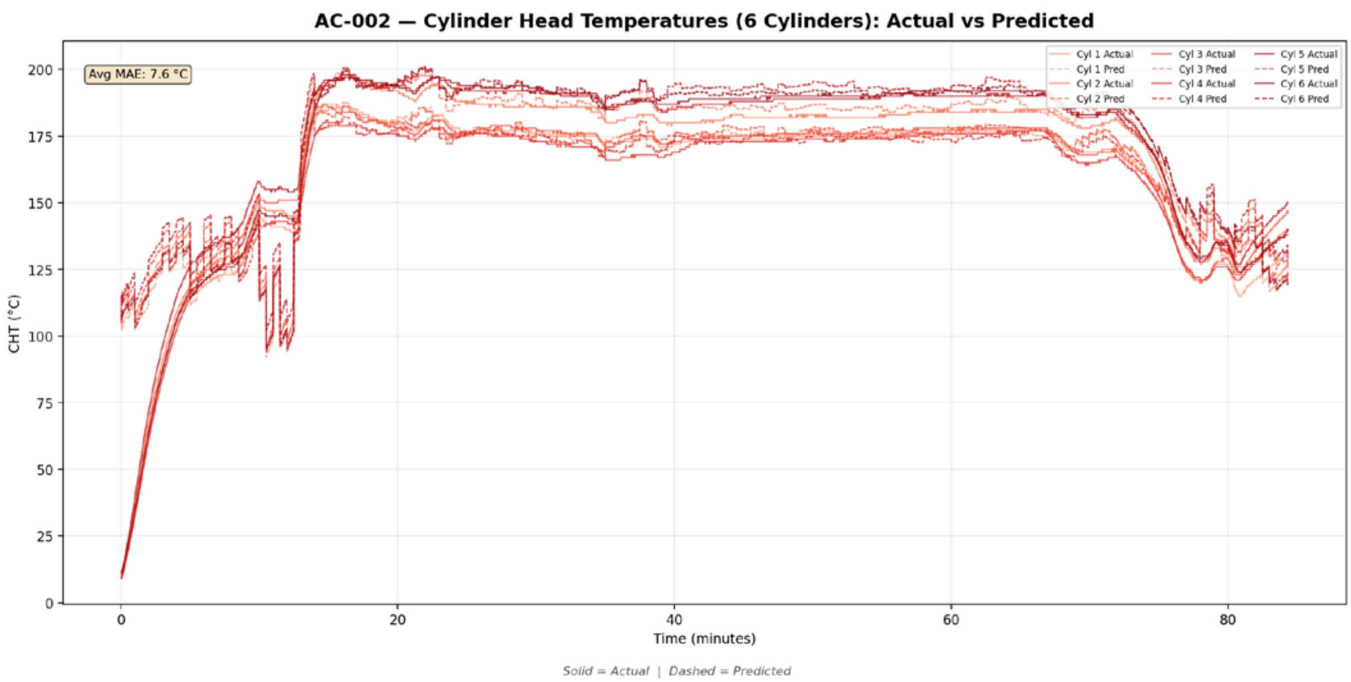


Figure 3: *Diffusion Model Predictions for Oil Pressure and Temperature, EGT, and CHT*

Minimum Viable Digital Twin

A functionally useful engine digital twin with EGT within 1-2%, CHT within 3-5% during stabilized cruise can be built from as few as 10–20 flights of real operational data.

With modern avionics logging at 1Hz, 20 flights of ~2 hours each yields ≈ 144,000 data points, achievable within the first month of operations for an actively flown aircraft.

Beyond 20 flights, improvement is real but marginal, with AC-001 demonstrating only 0.21 percentage points from the 20-flight elbow (1.9%) to full model saturation (1.69%).

Synthetic Data to Operational Intelligence

The synthetic data generated by the diffusion-TS digital twin enables a class of downstream capabilities that are impossible to build with real data alone. The core mechanism is the digital twin's prediction residuals, the difference between what the twin predicts and what the aircraft actually reports.

The validated digital twin architecture enables several downstream capabilities. The following describe operational applications currently in development.

Adaptive Alerting Thresholds

Static alerting thresholds ignore context. A CHT of 390°F during a high-power climb on a hot day may be entirely normal, while a CHT of 370°F during standard-day cruise may indicate a developing problem. Synthetic data enables context-aware adaptive thresholds by generating telemetry across the full matrix of flight phases, power settings, and environmental conditions.

What-If Scenario Testing

The X-Plane + diffusion model pipeline generates scenarios on demand that are too dangerous or expensive to produce with a real aircraft: sustained over-boost, progressive fuel contamination, slow alternator failure. This serves dual purposes: training the anomaly detector and validating that alerting behavior is appropriate.

Engine Characterization Map

Beyond inference of individual flight profiles, the diffusion-TS digital twin enables a capability unavailable from real flight data alone: a complete engine characterization map generated by systematically sweeping the aircraft's operating envelope. This map answers the question real data cannot: what does this engine do at every combination of power, altitude, temperature, and mixture, not just the combinations a pilot happened to fly.

The characterization map is generated by constructing synthetic input sequences that span the engine's full operating envelope and passing them through the trained diffusion model. For each operating point, a 120-row smooth input sequence is constructed (30-row linear ramp from idle to target conditions, followed by 90 rows of steady-state hold).

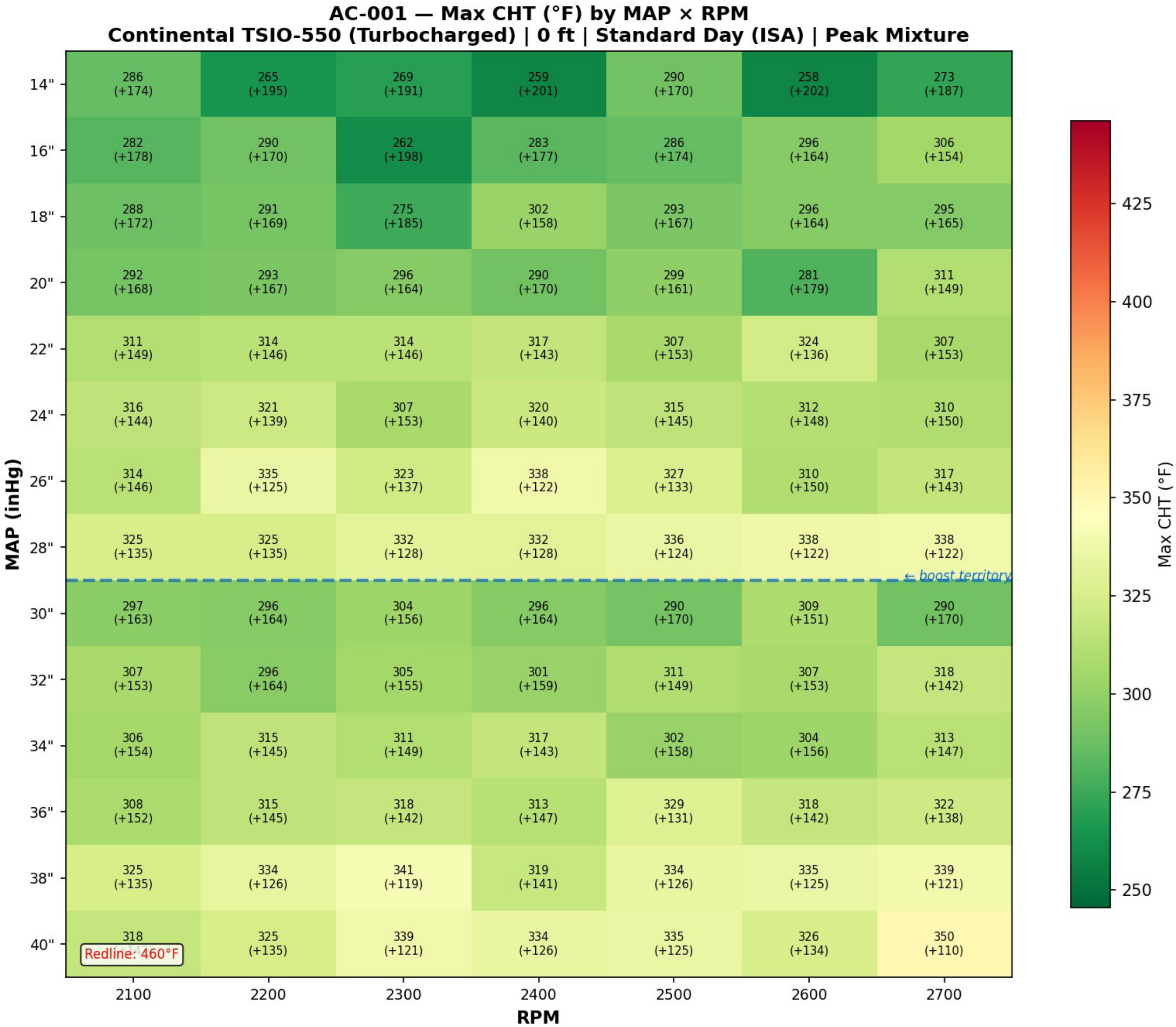


Figure 4: *Engine Characterization Map for CHT (°F) by Power Setting*

Figure 4 shows a CHT heatmap at 8,000 ft, standard day (ISA), peak mixture. Each cell shows max CHT and margin below the 460°F redline. High-MAP/high-RPM combinations (lower-right) carry only +94–+112°F of margin. Economy cruise settings (upper-left) carry 155–187°F. No real aircraft flew all 5,880 combinations; this map exists only because the digital twin can be queried at any operating point on demand.

The map produces actionable insights across five derived metrics: CHT Margin Heatmap (which power settings approach redline), Environmental Sensitivity (how much does a hot day compress thermal margins), Mixture Sensitivity (does the model recover the EGT peak at lean-of-peak), Cylinder Imbalance (which settings produce the greatest per-cylinder thermal spread), and Efficiency Map (where is best power-specific fuel consumption).

The most safety-critical operating conditions are systematically underrepresented in real flight data. High-power operations at high density altitude on hot days, conditions most likely to stress thermal margins, are precisely the conditions that careful operational flying minimizes. The diffusion-TS digital twin inverts this relationship. Because the model has learned the statistical structure of engine behavior across diverse operating conditions, it can extrapolate to the thermal margins.

Business Benefits

Safety and Certification

The digital twin's validated synthetic data generation changes the economics of AI-driven safety systems. Rather than waiting years for sufficient real anomaly data to accumulate across a fleet, operators can generate training data for specific degradation scenarios on demand: oil system deterioration, cylinder head thermal exceedances, electrical system fades. The ablation study establishes a concrete onboarding threshold: as few as 10–20 real flights produce a digital twin with EGT fidelity within 1–2% and CHT fidelity within 3–10%, depending on avionics quality. For an actively flown aircraft logging at 1Hz, 20 flights of approximately two hours each yield roughly 144,000 data points, achievable within the first month of operations.

This means new aircraft can generate operational anomaly intelligence within weeks of delivery, not the 12–18 months that fleet-average approaches typically require. For fleet operators, the implications are immediate: every aircraft in the fleet carries its own behavioral baseline from early in its service life, and anomaly detection begins long before conventional data accumulation would allow.

The certification implications are equally direct. Synthetic data generated by a validated per-asset digital twin provides a new class of evidence for safety validation, grounded in that aircraft's actual learned behavior rather than generic fleet models or physics-only simulations. Fleet managers gain consistent anomaly detection coverage across all aircraft regardless of age or accumulated flight hours, and safety margins become quantifiable: the engine characterization map demonstrated in this analysis maps thermal margin across 5,880 operating point combinations that no real aircraft has flown, identifying that high-MAP/high-RPM conditions carry as little as 94°F of margin below the 460°F CHT redline.

Reduced Maintenance Costs

By building per-asset digital twins rather than relying on fleet-average models, operators gain visibility into the unique behavior and wear patterns of each individual system. With AC-001's digital twin achieving 1.69% overall engine health MAPE at full training data, maintenance decisions shift from generalized thresholds and fixed schedules to precise, behavior-based intervention. When a specific aircraft's real telemetry deviates from its twin's predictions, that deviation carries diagnostic meaning; it reflects a change in that engine, not a statistical anomaly averaged across a fleet.

The training pipeline itself revealed an unexpected maintenance capability: model divergence as a diagnostic signal. During the ablation study, AC-002's oil temperature channel produced an anomalously high 17.4% MAPE, far outside the range explained by data volume or model architecture. Investigation confirmed a failing oil temperature sensor in active service, identified by the digital twin training process before conventional monitoring flagged it for repair. This transforms the training pipeline from a one-time model creation step into a continuous sensor health audit across the fleet.

The practical result is fewer unnecessary inspections, fewer missed faults, and longer component life. Operators inspect when the twin signals a behavioral change, not when a calendar interval expires or a fleet-wide threshold triggers. For a 200-aircraft regional fleet served by four R770 inference nodes, every aircraft in the fleet receives this level of individualized monitoring continuously.

Improved Uptime

Because each digital twin continuously learns how its specific asset behaves under real operating conditions, it detects subtle performance drift and emerging degradation long before traditional threshold-based alerting would trigger. The oil degradation scenario described earlier in this analysis illustrates the point: oil temperature rising incrementally, oil pressure dropping in small steps correlated with power settings, cylinder head temperatures trending upward; at any single point, every parameter remains within published normal ranges. A threshold-based system sees nothing. The digital twin, having learned the specific relationships between these parameters for this aircraft, flags the deviation.

This shifts operators from reactive repairs and calendar-based maintenance to predictive, condition-based intervention. Unplanned downtime drops because faults surface early, when they can be addressed during scheduled maintenance windows rather than as in-service failures. Cascading failures, where one undetected issue degrades a related system, become visible before they propagate. The operational result is higher asset utilization, more predictable fleet performance, and reduced disruption to revenue operations.

The dual-platform infrastructure architecture supports this directly. Training runs on GPU-dense PowerEdge XE7745 systems at a cadence the fleet requires, including initial model build, periodic retraining as the asset ages, or event-triggered retraining after maintenance actions. Inference runs continuously on PowerEdge R770 servers, processing incoming 1Hz telemetry from every monitored aircraft with anomaly flags delivered within seconds of data receipt. The infrastructure scales with the fleet: one R770 node for a 50-aircraft operator, four nodes for a 200-aircraft regional carrier, twenty nodes for a major carrier fleet of 1,000 aircraft.

Transferability Beyond Aviation

The generative digital twin architecture extends to any domain where safety-critical physical systems generate time-series telemetry and true anomaly data is rare. The closest adjacencies are in aerospace: eVTOL aircraft, autonomous air vehicles, and rotorcraft share the same telemetry-rich, failure-scarce data profile that makes per-asset digital twins valuable. Maritime propulsion, energy infrastructure, and industrial turbomachinery present structurally identical challenges: high-value assets, continuous sensor streams, and safety regimes where failures are too rare to train on and too costly to tolerate. Automotive powertrains and industrial IoT systems represent a broader application surface where the same closed-loop architecture applies: real telemetry trains the twin, the twin generates operational scenarios, deviations from the twin surface anomalies, and the twin improves as the asset operates.

Conclusion

This paper demonstrates that generative digital twins can move safety-critical AI from reactive monitoring to proactive operational intelligence. By training per-asset diffusion models on real telemetry, 550 Aero constructed high-fidelity digital twins that reproduce engine behavior within single-digit error rates during stabilized flight and extend visibility across operating conditions rarely captured in real data.

The systematic ablation study establishes a practical onboarding threshold: approximately 10–20 flights are sufficient to build a functionally useful digital twin, with model saturation reached near 60 flights. This defines a measurable path from asset delivery to operational intelligence within weeks, not years. The engine characterization methodology further expands the observable operating envelope beyond what real flights alone can provide, enabling operators to quantify safety margins, stress conditions, and efficiency trade-offs across thousands of operating points on demand.

At fleet scale, the training pipeline revealed a second, unexpected capability: model divergence itself becomes a diagnostic signal. The identification of a failing oil temperature sensor through abnormal MAPE performance demonstrates that digital twin training is not only model creation, but also continuous asset validation.

The dual-platform architecture validated at the Dell AI Innovation Lab, training on GPU-dense PowerEdge XE7745 systems and deploying inference at scale on PowerEdge R770 servers, demonstrates that generative digital twins are operationally practical. Training and inference workloads are matched to the appropriate infrastructure phase, enabling scalable fleet monitoring without requiring high-end hardware at every layer.

Digital twins are not simply a modeling exercise; they represent a shift in how physical systems are governed. Organizations operating fleets, factories, or critical infrastructure should begin by capturing high-quality telemetry, defining per-asset modeling strategies, and aligning infrastructure to the digital twin lifecycle. The technical barrier is lower than assumed. The operational upside is immediate: safer systems, higher uptime, and lower maintenance cost.

Acknowledgments

The authors would like to thank 550 Aero for demonstrating their work and Dell Technologies for providing access to the PowerEdge R770 and XE7745 hardware platforms and technical expertise. Special recognition goes to the engineering teams at Broadcom for their collaboration on network optimization strategies.

Important Information About this Report

CONTRIBUTORS

Brian Martin
AI Data Center Performance | Signal65

PUBLISHER

Ryan Shrout
President and GM | Signal65

INQUIRIES

Contact us if you would like to discuss this report and Signal65 will respond promptly.

CITATIONS

This paper can be cited by accredited press and analysts, but must be cited in-context, displaying author's name, author's title, and "Signal65." Non-press and non-analysts must receive prior written permission by Signal65 for any citations.

LICENSING

DISCLOSURES

Signal65 provides research, analysis, advising, and consulting to many high-tech companies, including those mentioned in this paper. No employees at the firm hold any equity positions with any companies cited in this document.

ABOUT SIGNAL65

Signal65 is a leading research organization specializing in enterprise AI infrastructure optimization and deployment strategies. Our lab focuses on evaluating and optimizing AI hardware and software solutions for real-world enterprise applications, with particular expertise in large language models, retrieval-augmented generation systems, and distributed AI architectures.

For more information, visit signal65.com
or contact research@signal65.com

IN PARTNERSHIP WITH

Dell Technologies

550 AERO

CONTACT INFORMATION
Signal65 | signal65.com